SEMICONDUCTOR STRUCTURES WITH STRUCTURAL HOMOGENEITY

Related Applications

This application claims the benefit of U.S. Provisional Application Serial Number 60/442,784, filed on January 27, 2003, the entire disclosure of which is hereby incorporated by reference.

Field of the Invention

This invention relates generally to semiconductor substrates and particularly to substrates with strained semiconductor layers.

Background

"Virtual substrates" based on silicon (Si) and germanium (Ge) provide a platform for new generations of very large scale integration (VLSI) devices that exhibit enhanced performance in comparison to devices fabricated on bulk Si substrates. The important component of a SiGe virtual substrate is a layer of SiGe that has been relaxed to its equilibrium lattice constant (i.e., one that is larger than that of Si). This relaxed SiGe layer may be directly applied to a Si substrate (e.g., by wafer bonding or direct epitaxy), or atop a relaxed graded SiGe buffer layer in which the lattice constant of the SiGe material has been increased gradually over the thickness of the layer. The SiGe virtual substrate may also incorporate buried insulating layers, in the manner of a silicon-on-insulator (SOI) wafer. To fabricate high-performance devices on these platforms, thin strained layers of semiconductors, such as Si, Ge, or SiGe, are grown on the relaxed SiGe virtual substrates. The resulting biaxial tensile or compressive strain alters the carrier mobilities in the layers, enabling the fabrication of high-speed and/or low-power-consumption devices. The thin strained semiconductor layers may also be subsequently transferred to other substrates having insulator layers by methods such as wafer bonding, thus creating strained-semiconductor-on-insulator (SSOI) wafers.

In certain cases the microstructure of semiconductor graded buffer layers as grown may be less than ideal depending on the growth conditions. For example, SiGe buffer layers

deposited at temperatures below 850 °C may not attain the relaxation state desired for strained Si applications, i.e., >98%. In addition, the density of threading dislocations may be higher than desired. Furthermore, both high and low temperature growth conditions may result in as-grown graded buffer layers having top surfaces that are rougher than the ultra-planar surfaces preferable for growth of relaxed semiconductor cap layers with subsequent strained semiconductor layer deposition (e.g., regrowth of SiGe layers containing 20% Ge, followed by deposition of strained Si). This roughness may carry over and increase in subsequently formed layers. In addition, roughness on a layer surface negatively impacts the ability of laser scanning tools to perform optical inspection for defects in the layer before and after planarization and regrowth. Roughness appears in the scattered signal of the laser scanner as an elevated level of "haze" or background noise, reducing the ability of the tool to detect small defects in and on the layer. It is desirable, therefore, to reduce this roughness in semiconductor layers.

Summary

One technique suitable for fabricating strained Si wafers may include the following steps:

1. Providing a silicon substrate;
2. Epitaxially depositing a relaxed, graded SiGe buffer layer to some final Ge composition on the silicon substrate;
3. Epitaxially depositing a relaxed SiGe cap layer having a constant composition on the SiGe buffer layer;
4. Annealing the layers at a temperature greater than a growth temperature of the layers to relax strain or modify the morphology of the layers, at any point during or after Steps 2 and 3;
5. Planarizing a surface of the SiGe cap layer by, e.g., chemical mechanical polishing (CMP), and cleaning the resulting planarized surface;
6. Epitaxially depositing a relaxed SiGe regrowth layer having a constant composition on the planarized surface; and
7. Epitaxially depositing a strained Si layer on the SiGe regrowth layer.
8. Measuring the surface quality of the strained Si layer using laser scanning techniques.

Annealing at elevated temperatures may improve the properties of layers deposited at relatively low temperatures, e.g., below 850 °C. Various layer properties, in addition to

relaxation and threading dislocation densities, are important for making strained semiconductor layers, e.g., strained silicon layers. For example, at high temperature growth conditions (>850 °C), graded and constant composition SiGe buffer layers may contain microstructural phenomena such as decomposition. Decomposition may sometimes be observed as narrow vertical bands of varying composition, i.e., vertical superlattices.

Elevated temperature annealing before, after, or between planarization process steps may be used to improve the microstructure of semiconductor layers. Compositional variation within layers is reduced, thereby enabling the formation of layers with top surfaces that remain smooth even after cleaning steps that etch different compositions at different rates.

In some embodiments, compositional superlattices may be avoided by appropriate selection of semiconductor layer growth parameters and regrowth layer parameters.

In an aspect, the invention features a method for forming a semiconductor structure, the method including providing a substrate, and forming a semiconductor layer over a top surface of the substrate, the semiconductor layer including at least two elements, the elements being distributed to define an initial compositional variation within the semiconductor layer. The semiconductor layer is annealed to reduce the initial compositional variation.

One or more of the following features may be included. The substrate may have a first lattice constant, the semiconductor layer may have a second lattice constant, and the first lattice constant may differ from the second lattice constant. The first element may have a first concentration, a second element may have a second concentration, and each of the first and second concentrations may be at least 5%. The initial compositional variation may vary periodically within the semiconductor layer in a direction perpendicular to a semiconductor layer deposition direction. The compositional variation may define a column within the semiconductor layer, the column having a width and a period. The columnar period may be less than approximately 2000 nanometers (nm), e.g., less than approximately 1000 nm.

The semiconductor layer may be annealed at an annealing temperature and/or for a duration sufficient to diffuse at least one of the two elements through a diffusion length at least equal to a quarter of the columnar period.

The initial compositional variation may vary in a direction parallel to a semiconductor layer deposition direction and define a superlattice having a periodicity. The superlattice periodicity may be less than approximately 100 nm, preferably less than approximately 50 nm,

and more preferably less than approximately 10 nm. The semiconductor layer may be annealed at an annealing temperature sufficient to diffuse at least one of the two elements through a diffusion length at least equal to a quarter-period of the superlattice and/or for a duration sufficient to diffuse at least one of the two elements through a diffusion length at least equal to a quarter-period of the superlattice.

The semiconductor layer may be annealed at an annealing temperature greater than the deposition temperature. The annealing temperature may be greater than about 800 °C, e.g., greater than about 1000 °C.

The semiconductor layer may be annealed at an annealing temperature below a melting point of the semiconductor layer, e.g., less than about 1270°C.

At least one of the at least two elements may be silicon and/or germanium. A top surface of the semiconductor layer may be planarized. The top surface of the semiconductor layer may be planarized before, while, or after the semiconductor layer is annealed. Planarizing may include chemical-mechanical polishing, plasma planarization, wet chemical etching, gas-phase chemical etching [preferably at elevated temperature, e.g., above 900°C, in an ambient including an etch species, e. g., hydrogen chloride (HCl)], oxidation followed by stripping, and/or cluster ion beam planarization.

Chemical-mechanical polishing may include a first and a second step and the semiconductor layer may be annealed between the first and the second chemical-mechanical polishing steps and/or before the first chemical-mechanical polishing step. The planarization may include a high temperature step and the semiconductor layer may be annealed during the high temperature planarization step.

A top surface of the semiconductor layer may be bonded to a wafer, and at least a portion of the substrate may be removed, such that at least a portion of the semiconductor layer remains bonded to the wafer after the portion of the substrate is removed.

A second layer may be formed over the semiconductor layer subsequent to planarizing the top surface of the semiconductor layer. The second layer may include a material having a lattice constant substantially equal to or substantially different from a lattice constant of the semiconductor layer. A top surface of the second layer may be bonded to a wafer and at least a portion of the substrate may be removed, such that at least a portion of the second layer remains bonded to the wafer after the portion of the substrate is removed.

A second layer may be formed over the semiconductor layer subsequent to planarizing the top surface of the semiconductor layer. The second layer may include a material having a lattice constant substantially equal to or substantially different from a lattice constant of the semiconductor layer. A top surface of the second layer may be bonded to a wafer, and at least a portion of the substrate may be removed, with at least a portion of the second layer remaining bonded to the wafer after the portion of the substrate is removed. The second layer may include (i) a lower portion having a superlattice and (ii) an upper portion disposed over the lower portion, the upper portion being substantially free of a superlattice.

The semiconductor layer may have an undulating surface. The undulating surface may be formed during deposition of the semiconductor layer. The substrate may have an undulating substrate surface, and the undulating substrate surface induces the formation of the undulating surface of the semiconductor layer. The undulating surface may have an amplitude, the initial compositional variation may define a superlattice having a periodicity, and the periodicity of the superlattice may be less than the amplitude of the undulating surface.

A relaxed graded layer may be formed over the substrate, such that the semiconductor layer is formed over the relaxed graded layer. The relaxed graded layer may serve to provide the semiconductor layer with a lattice spacing different from that of the substrate while reducing defect nucleation. A protective layer may be formed over the semiconductor layer prior to annealing the semiconductor layer. The protective layer may include a material that is substantially inert with respect to the semiconductor layer, such as, for example, silicon dioxide or silicon nitride. The anneal may be performed as a batch process on multiple wafers at once, for example, in a tube furnace, to improve throughput and economics.

In another aspect, the invention features a method for forming a semiconductor structure, including providing a substrate, and selecting a first plurality of parameters suitable for forming a semiconductor layer over a top surface of the substrate, the semiconductor layer including at least two elements, the elements being distributed to define a compositional variation within the semiconductor layer. The semiconductor layer having a haze is formed, and the semiconductor layer is planarized to remove the haze.

One or more of the following features may be included. Forming the semiconductor layer may include forming a lower portion having a superlattice, and forming an upper portion over the lower portion, the upper portion being substantially free of a superlattice. The first

plurality of parameters may include temperature, precursor, growth rate, and/or pressure. The semiconductor layer may be cleaned after planarizing, withthe semiconductor layer remaining substantially haze-free after cleaning. A second plurality of parameters may be selected that is suitable for forming a substantially haze-free regrowth layer over the semiconductor layer, the semiconductor layer including at least two elements, the elements being distributed to define a compositional variation within the semiconductor layer. The substantially haze-free regrowth layer may be formed. The first plurality of parameters may include a first temperature, the second plurality of parameters may include a second temperature, and the first temperature may be higher than the second temperature. The first plurality of parameters include a first growth rate, the second plurality of parameters may include a second growth rate, and the first growth rate may be higher than the second growth rate. Forming the regrowth layer may include forming a lower portion having a superlattice and forming an upper portion over the lower portion, the upper portion being substantially free of a superlattice.

In another aspect, the invention features a semiconductor structure including a substrate, and a semiconductor layer disposed over the substrate, the semiconductor layer including at least two elements and having a top surface. The semiconductor layer top surface is substantially haze-free.

One or more of the following features may be included. A portion of the semiconductor layer disposed below the top surface may include a superlattice. A relaxed graded layer may be disposed between the substrate and the semiconductor layer. The semiconductor layer top surface may have a roughness root-mean-square of less than 10 angstroms (Å), preferably less than 5 Å in a scan area of 40 μm × 40 μm, and a contamination level of less than 0.29 particles/cm^2, the particles having a diameter greater than 0.12 micrometers (μm). Preferably, the roughness is less than 1 Å root-mean-square in a scan area of 1 μm × 1 μm.

The semiconductor layer top surface may have a roughness of less than 10 Å, preferably less than 5 Å root-mean-square in a scan area of 40 μm × 40 μm and a contamination level of less than 0.16 particles/cm^2, the particles having a diameter greater than 0.16 μm. Preferably, the roughness is less than 1 Å root-mean-square in a scan area of 1 μm × 1 μm.

The semiconductor layer top surface may have a roughness of less than 10 Å, preferably less than 5 Å root-mean-square in a scan area of 40 μm × 40 μm and a contamination level of

less than 0.08 particles/cm^2, the particles having a diameter greater than 0.2 μm. Preferably, the roughness is less than 1 Å root-mean-square in a scan area of 1 μm × 1 μm.

The semiconductor top surface may have a roughness of less than 10 Å, preferably less than 5 Å root-mean-square in a scan area of 40 μm × 40 μm and a contamination level of less than 0.019 particles/cm^2, the particles having a diameter greater than 1 μm. Preferably, the roughness is less than 1 Å root-mean-square in a scan area of 1 μm × 1 μm.

The semiconductor layer top surface may have a roughness of less than 0.5 Å root-mean-square in a scan area of 1 μm × 1 μm and a contamination level of less than 0.09 particles/cm^2, the particles having a diameter greater than 0.09 μm.

In another aspect, the invention features a semiconductor structure including a substrate, and a semiconductor layer disposed over the substrate, the semiconductor layer including at least two elements. A regrowth layer is disposed over the semiconductor layer, the regrowth layer having a top surface that is substantially haze-free.

One or more of the following features may be included. The regrowth layer may include a semiconductor material, such as silicon. The regrowth layer may be strained. A portion of the regrowth layer disposed below the regrowth layer top surface may include a superlattice.

In another aspect, the invention features a semiconductor structure including a wafer, and a semiconductor layer bonded to the wafer, the semiconductor layer having a top surface that is substantially haze-free.

One or more of the following features may be included. The semiconductor layer may include silicon and/or germanium. The semiconductor layer may be strained. The wafer may include an insulating layer. The insulating layer may include silicon dioxide.

Brief Description of the Drawings

Figures 1 - 6 and 8 - 11 are schematic cross-sectional views of semiconductor substrates illustrating superlattices, columnar structures, and processes for forming homogeneous, smooth semiconductor layers; and

Figure 7 is a diagram illustrating the temperature and time dependence of diffusion of Ge in Si.

Detailed Description

Roughness on semiconductor graded buffer layers may be separated into two components, each with distinct characteristics. A first component is a cross-hatch that arises from strain fields created by the formation of misfit dislocations. Cross-hatch has the form of a network of perpendicular waves with several characteristic wavelengths. For many graded buffer layers formed on wafers, for example layers constituted of group IV or III-V semiconductors with diamond cubic or zinc blende crystal structures, this cross-hatch is generally oriented in the <110> in-plane direction of the wafers. This relatively widely spaced component of surface texture may be likened to a surface feature referred to in the SEMI Specifications as "waviness." A second component, present in different degrees depending on the growth conditions, is small-scale roughness with no obvious directionality, a smaller amplitude, and a shorter spatial wavelength than the cross-hatch. This fine-scale roughness may be a major contributor to haze measured on semiconductor layers by laser defect scanning tools. Methods for reducing or eliminating both cross-hatch and fine scale roughness are described below.

Referring to Figure 1, an epitaxial wafer 8 has a plurality of layers 10 disposed over a substrate 12. Substrate 12 may be formed of a semiconductor, such as Si, Ge, or SiGe. Substrate 12 may also include an insulator layer (not shown). The plurality of layers 10 formed on a top surface 13 of substrate 12 includes a graded buffer layer 14, which may be relaxed and may be formed of $Si_{1-y}Ge_y$, with a maximum Ge content of, e.g., 10 - 100% (i.e., y = 0.1 - 1.0) and a thickness T_1 of, for example, greater than or equal to 0.5 μm, e.g., 0.5 - 10 μm. A semiconductor layer 16 is disposed over graded buffer layer 14. Semiconductor layer 16 may be relaxed, and may contain at least two elements. The substrate may have a first lattice constant and the semiconductor layer 16 may have a second lattice constant, such that the first lattice constant differs from the second lattice constant. The first element may have a first concentration and the second element may have a second concentration, and each of the first and second concentrations may be greater than 5%. The two elements may be, for example, silicon and germanium (e.g., $Si_{1-x}Ge_x$). $Si_{1-x}Ge_x$ may have a Ge content of, for example, 10 - 100 % (i.e., x = 0.1 - 1.0), and a thickness T_2 of, for example, 0.2 - 2 μm. In some embodiments, $Si_{1-x}Ge_x$ may include $Si_{0.80}Ge_{0.20}$ and T_2 may be approximately 1.5 μm. Semiconductor layer 16 may be >90% relaxed, as determined by triple axis x-ray diffraction, and may have a threading dislocation

density of <1 x 10^6 cm^{-2}, as determined by etch pit density (EPD) and plan-view transmission electron microscopy (PVTEM) analysis.

Graded layer 14 and semiconductor layer 16 may be formed by epitaxy, such as by atmospheric-pressure chemical vapor deposition (APCVD), low- (or reduced-) pressure CVD (LPCVD), ultra-high-vacuum CVD (UHVCVD), or by molecular beam epitaxy (MBE). The epitaxial deposition system may be a single-wafer or multiple-wafer batch reactor. The growth system may include a horizontal flow reactor, in which process gases are introduced into the reactor from one side and exit the reactor from another side, after passing over one or more substrates. The growth system may also utilize a low-energy plasma to enhance layer growth kinetics. The deposition temperature may be 500 - 1200 °C.

Substrate 12, graded layer 14, and semiconductor layer 16 may be formed from various materials systems, including various combinations of group II, group III, group IV, group V, and group VI elements. For example, each of substrate 12, graded layer 14, and semiconductor layer 16 may include a III-V compound. Substrate 12 may include gallium arsenide (GaAs), and graded layer 14 and semiconductor layer 16 may include indium gallium arsenide (InGaAs) or aluminum gallium arsenide (AlGaAs). These examples are merely illustrative, and many other material systems are suitable.

In alternative embodiments, semiconductor layer 16 is tensilely strained (e.g., Si_xGe_{1-x} disposed over Si_yGe_{1-y} where $y < x$). In other embodiments, semiconductor layer 16 is compressively strained (e.g., Si_xGe_{1-x} disposed over Si_yGe_{1-y} where $y > x$). In these cases, semiconductor layer 16 may be disposed over a relaxed semiconductor layer. In some embodiments, a strained layer (not shown) may be formed on a top surface of semiconductor layer 16 or graded layer 14.

Referring to Figure 2 as well as to Figure 1, as deposited, a distribution of the elements from which semiconductor layer 16 is formed may have an initial compositional variation 20. For example, if semiconductor layer 16 includes 20% Ge ($Si_{0.80}Ge_{0.20}$), the actual Ge concentration within layer 16 may vary by a total of 4%, e.g., 18 - 22%. This initial compositional variation 20 may vary in semiconductor layer 16 in a direction parallel to a deposition direction 22 thereof.

Compositional variation 20 may define a superlattice 24 having a periodicity P_1. Superlattice 24 has alternating regions with low 28 and high 29 concentrations of an element,

e.g., Ge, alternating in the same layer, such as in semiconductor layer 16. Such alternation may occur in a horizontal flow deposition reactor, in which a higher fraction of an element is incorporated at a leading edge of a substrate, i.e., an edge of wafer 8. The element fraction, e.g., Ge concentration, may alternate vertically within semiconductor layer 16 because substrate 10 may be rotated during deposition, thus changing the leading edge first exposed to gas flow. Depending on deposition parameters, alternating compositions within a layer may also occur in layers formed in other types of deposition systems. Superlattice 24 may have a superlattice periodicity P_1. Periodicity P_1 may be less than approximately 100 nm, including less than 50 nm or less than 10 nm. In an embodiment, periodicity P_1 may be 8 nm with, e.g., region 29 having a thickness of 4 nm with Ge concentration above, e.g., 20% and region 28 having a thickness of 4 nm with Ge concentration below, e.g., 20%.

Referring to Figure 3, semiconductor layer 16 may be formed over graded buffer layer 14 having a top surface 15 that may not be completely smooth, i.e., it may have cross-hatch formed by strain fields arising from the formation of misfit dislocations. A cross-hatch may have, for example, a relatively high Ge concentration at a peak and a relatively low Ge concentration in a trough. Cross-hatch may have a wavelength of 1 - 10 μm and an amplitude of 1 - 100 nm. Graded buffer layer surface 15 may also have fine-scale roughness, with a wavelength of, e.g., 10 - 100 nm and a height of 1 - 50 Å. Both cross-hatch and fine-scale roughness may carry over to cause undulation 30 in a top surface 32 of semiconductor layer 16. Undulation 30 may be formed during deposition of semiconductor layer 16. Undulation 30 has an amplitude A that may be greater than periodicity P_1 of superlattice 24.

Referring to Figure 4, semiconductor layer surface 32 may be planarized by, e.g., CMP. Planarization exposes lateral composition variations on planarized semiconductor layer surface 32. The periodicity P_1 of elements disposed in semiconductor layer 16, i.e., superlattice 24, however, may cause problems with subsequent processing. For example, maintaining the planarity of semiconductor layer 16 may be challenging. Cleaning steps after planarization may re-roughen surface 32. A wet cleaning solution whose removal rate is compositionally dependent may result in a rough top surface if there is lateral compositional variation in the layer being cleaned and the removal rate is compositionally dependent. Such a solution may, for example, selectively etch portions of layer 16 with higher concentrations of a particular element, such as region 29 having a higher concentration of, e.g., Ge, more quickly than portions of layer

16 with a lower concentration of the same element, such as region 28 having a lower concentration of, e.g., Ge. An example of such a wet etch is RCA SC1, i.e., ammonium hydroxide, hydrogen peroxide, and deionized water at a ratio of, e.g., 1:1:10 or 1:1:100, at 40 - 80 °C for about 10 minutes, with or without megasonic agitation.

Referring to Figure 5, in an alternative embodiment, initial compositional variation 20 may vary in semiconductor layer 16 in a direction perpendicular to the deposition direction 22 thereof. Initial compositional variation 20 may define a column 50 within semiconductor layer 16. Column 50 may have an irregular cross-section. Column 50 may form as a result of an interaction between the cross-hatch formed on surface 15 of graded layer 14 and superlattice 24 (see, e.g., Figures 2 - 4). This interaction may cause decomposition during the formation of semiconductor layer 16, resulting in the formation of a plurality of columns 50 having a relatively high concentration of an element, e.g., Ge, alternating with a plurality of columns 52 having a relatively low concentration of the same element. Column 50 and column 52 may each have a width W_1 less than approximately 1000 nm, such that columnar period P_2 , including column 50 and column 52 (one dark region and one light region in Figure 5) is less than approximately 2000 nm. In some embodiments, columnar period P_2 may be less than 1000 nm. Semiconductor layer surface 32 may be planarized, e.g., by CMP. The presence of columns 50, 52 with varying compositions in semiconductor layer 16, however, may cause problems with subsequent processing. For example, maintaining the planarity of semiconductor layer 16 may be challenging. Cleaning steps after planarization may re-roughen surface 32. Cleaning solutions, such as RCA SC1 may selectively etch faster portions of layer 16 with higher concentrations of a particular element, such as columns 50 having a higher concentration of, e.g., Ge than portions of layer 16 with lower concentrations of the same element, such as columns 52 having a lower concentration of, e.g., Ge.

Referring to Figure 6, the initial compositional variation within semiconductor layer 16 may be reduced by annealing semiconductor layer 16. The resulting reduction of the initial compositional variation may substantially eliminate superlattice 24, as well as columns 50, 52, resulting in a relatively homogeneous compositional distribution within semiconductor layer 16. The relatively uniform composition of semiconductor layer 16 reduces the aforementioned effects of cleaning steps, i.e., non-uniform etch rates of semiconductor layer 16 regions with varying compositions, resulting in roughening of semiconductor layer surface 32. Annealing

may increase the amplitude and wavelength of the cross-hatch, but reduces the short wavelength roughness. The cross-hatch may have a wavelength sufficiently long so that a small increase in the long wavelength roughness (>1 μm) may not affect optical scanning measurements of semiconductor layer surface 32.

Referring to Figures 3 and 6, the annealing temperature may be sufficient to diffuse at least one of the at least two elements included in semiconductor layer 16 through a diffusion length at least equal to one-quarter the period P_l of superlattice 24, in a cost-effective time. For example, to diffuse Ge through a diffusion length of 100 nm, the annealing temperature may be at least 850°C at a duration of 300,000 seconds (83.3 hours). This temperature and duration may be derived from the following equations. The diffusion length x may be calculated by:

$$x = 2*(Dt)^{0.5} \qquad \text{(Equation 1)}$$

where

x is the characteristic diffusion length,

D is the characteristic diffusion coefficient of one of the at least two elements in another of the at least two elements, and

t is the diffusion time.

The diffusion coefficient D is given by the following:

$$D = D_o \exp(-E/kT) \qquad \text{(Equation 2)}$$

where

D_o is the pre-exponential factor,

E is the activation energy,

k is the Boltzmann constant, and

T is the annealing temperature (in degrees Kelvin).

For example, for germanium diffusing in silicon, the following values may be obtained from published literature: $D_o = 6.26 \times 10^5$ cm²/sec, E = 5.28 eV, and $k = 8.63 \times 10^{-5}$ eV/K. Using these values, the characteristic diffusion distance may be calculated for a range of anneal times, and plotted versus temperature (see, e.g., Figure 7). The various values of the diffusion constants for germanium in silicon that are available may produce somewhat different results (see below). In some embodiments, the duration of the anneal is selected to be sufficient to

diffuse at least one of the at least two elements included in semiconductor layer 16 through a diffusion length at least equal to a quarter of the period P_1 of superlattice 24, at an acceptable temperature, i.e., a temperature high enough to provide adequate throughput without damaging the substrate or melting semiconductor layer 16. This temperature may be greater than about 800 °C and less than about 1270 °C. For example, to diffuse Ge through a diffusion length of at least 100 nm, the duration of the annealing may be at least 12 seconds at a temperature of 1250°C. This duration may be derived from equations 1 - 2 and/or Figure 7. Referring to Figures 5 and 6, the annealing temperature may be sufficient to diffuse one or more of the elements included in semiconductor layer 16 through a diffusion length at least equal to a quarter of the columnar period P_2 (in an economically acceptable time). For example, to diffuse Ge through a diffusion length of at least 1000 nm, the annealing temperature may be at least 1050°C at a duration of 300,000 sec (83.3 hours). The appropriate annealing temperature may be derived from the equations 1 - 2 above or Figure 7. In some embodiments, the duration of the anneal may be selected to be sufficient to diffuse at least one of the at least two elements included in semiconductor layer 16 through a diffusion length at least equal to a quarter of the columnar period P_2 . For example, to diffuse Ge through a diffusion length of at least 1000 nm, the duration of the annealing may be at least 1200 sec (20 minutes) at a temperature of 1250°C. This duration may be derived from equations 1 - 2 and/or Figure 7.

Referring to Figures 3, 5, and 6, in some embodiments, semiconductor layer 16 is annealed at an annealing temperature greater than a deposition temperature of semiconductor layer 16. For example, the annealing temperature may be greater than about 800 °C, or greater than about 1000 °C. The annealing temperature may also be less than a melting point of semiconductor layer 16. For example, for semiconductor layer 16 including $Si_{0.8}Ge_{0.2}$, the annealing temperature may be less than about 1270 °C. A dislocation density in semiconductor layer 16 may remain substantially unchanged during the annealing step.

Referring to Figure 6, after an annealing step, semiconductor layer 16 has a relatively homogeneous compositional distribution. Top surface 32 of semiconductor layer 16 may be planarized. This planarization may be performed before, during, or after the annealing step. Planarization may be performed by one of several methods, including CMP, plasma planarization, wet chemical etching, gas-phase chemical etching (preferably at elevated temperature, e.g., above 900°C, in an ambient including an etch species, e. g., HCl), oxidation

followed by stripping, and cluster ion beam planarization. In some embodiments, CMP includes a first (stock) and a second (final) step. The stock polish removes a larger fraction of the total amount of material to be removed (~0.5 μm) and leaves a semi-polished surface. The final polish step removes a smaller fraction of the total amount of material to be removed (< 0.1 microns) and produces a smooth polished surface. Semiconductor layer 16 may be annealed before or after the first CMP step. The anneal step may provide a greater benefit in terms of layer homogenization, but at perhaps higher cost, if it is inserted between two steps of the planarization process, e.g., between the stock and final polishing steps. The removal of the cross-hatch by the stock polish step before the anneal step may allow the threading dislocations to move more freely to the wafer edge during the anneal. Performing the final polish step after the anneal may be preferable for obtaining a smooth surface for the regrowth process (see, e.g., Figure 7). The anneal may be performed as a batch process on multiple wafers at once, for example, in a tube furnace, to improve throughput and economics.

Referring to Figure 8 as well as to Figure 7, after planarization, top surface 32 of semiconductor layer 16 may be bonded to a wafer 40. Subsequently, at least a portion of substrate 12 may be removed by, e.g., a wet etch step or a delamination process. After the removal of at least the portion of substrate 12, at least a portion of semiconductor layer 16 remains bonded to the wafer 40. In an embodiment, all of substrate 12 may be removed, and the semiconductor layer 16 may have a second substantially haze-free top surface 42. Second top surface 42 may be planarized (i.e., smoothed) after removal of substrate 12. Planarizing may include chemical-mechanical polishing, plasma planarization, wet chemical etching, gas-phase chemical etching (preferably at elevated temperature, e.g., above 900°C, in an ambient including an etch species, e. g., HCl), oxidation followed by stripping, and/or cluster ion beam planarization. Wafer 40 may include a second substrate 42 formed of a semiconductor, such as Si, Ge, or SiGe. Second substrate 42 may also be formed of an insulating material such as sapphire (Al_2O_3) or glass. Wafer 40 may also include an insulating layer 44 disposed over substrate 42 and formed from, e.g., silicon dioxide. This process may be used to, e.g., prepare a semiconductor-on-insulator (SOI) substrate or an SSOI substrate.

Referring to Figures 7 and 9, after planarization of top surface 32 of semiconductor layer 16, a second layer 50 may be formed over semiconductor layer 16. Second layer 50 may include, e.g., a semiconductor material including at least one of a group II, a group III, a group

IV, a group V, and a group VI element, and may be formed by, e.g., CVD. Second layer 50 may have a lattice constant substantially equal to a lattice constant of semiconductor layer 16. Second layer 50 may also be a regrowth layer formed from the same material as semiconductor layer 16. Alternatively, the lattice constant of second layer 50 may be substantially different from the lattice constant of semiconductor layer 16. The lattice constant of second layer 50 may be less than that of semiconductor layer 16, in which case second layer 50 may be tensilely strained. For example, semiconductor layer 16 may include $Si_{1-x}Ge_x$ and second layer 50 may include $Si_{1-z}Ge_z$, with $z < x$. In another embodiment, the lattice constant of second layer 50 may be greater than the lattice constant of semiconductor layer 16, in which case second layer 50 will be compressively strained. For example, semiconductor layer 16 may include $Si_{1-x}Ge_x$ and second layer 50 may include $Si_{1-z}Ge_z$, with $z > x$. A top surface of second layer 50 may be bonded to wafer 40. Subsequently, at least a portion of substrate 12 may be removed by, e.g., a wet etch step or a delamination process. After the removal of at least the portion of substrate 12, at least a portion of the second layer 50 remains bonded to wafer 40. This process may be used to, e.g., prepare a SOI substrate or a SSOI substrate.

Referring again to Figure 6, after annealing and planarization, top surface 32 of semiconductor layer 16 is substantially haze-free. Haze is caused by background scattering of a surface, and is directly proportional to the roughness of the surface. Surface roughness may include features on several different spatial wavelengths. The cross-hatch features may typically be several micrometers (e.g., 1 μm -10 μm) in wavelength, while a fine-scale roughness may also be present on a shorter length scale (<1 μm). Surface roughness may be measured by atomic force microscopy (AFM), with a tool like the Dimension 3100 from Veeco Instruments, Woodbury, New York.) Haze may be measured by a light-scattering tool, such as various models of the SURFSCAN tool manufactured by KLA-Tencor, San Jose, California or the Film Inspection Tool (FIT) / Advanced Wafer Inspection System (AWIS) manufactured by ADE Corporation, Westwood, Massachusetts. In such laser-based particle or defect detection systems for semiconductor wafers, surface roughness causes may cause light scattering, which is termed "haze." The optical architecture of the system, i.e., the wavelength of the laser, the incident beam angle, and the polar and azimuthal angles of the collection detector(s) determines the spatial wavelengths of roughness to which the system is sensitive. For example, the SURFSCAN 6220, SURFSCAN SP1-TBI dark-field narrow channel with normal incidence

beam (DNN), and ADE FIT/AWIS front channel are sensitive primarily to surface roughness features with a wavelength of ~1-10 microns, which corresponds to the cross-hatch feature. In contrast, the SURFSCAN SP1-TBI dark-field wide channel with normal incidence (DWN), the ADE FIT/AWIS back and center channels, and SURFSCAN SP1 dark-field narrow channel with oblique incidence (DNO) are primarily sensitive to surface features with a spatial wavelength of <1 μm, which corresponds to fine-scale roughness. Lower haze values indicate smoother (lower roughness) surfaces, which are generally preferred. Haze values measured by a SURFSCAN 6220 for a high-quality surface are preferably less than 20 parts per million (ppm), more preferably less than 5 ppm, and most preferably less than 1 ppm. Haze values measured by the ADE FIT/AWIS back and center channels or by the SURFSCAN SP1-TBI DNO channel are preferably less than 0.2 ppm and more preferably less than 0.05 ppm.

By annealing semiconductor layer 16, the compositional variation is homogenized. This uniform composition enables the planarization of top surface 32, as well as cleaning of top surface 32, without the re-introduction of roughness. Top surface 32 of semiconductor layer 16 may, therefore, be both smooth and clean. For example, top surface 32 may have a roughness root-mean-square (RMS) of less than 5 Å (in a 40 μm × 40 μm scan area), less than 1 Å (in a 1 μm × 1 μm scan area) and a contamination level of less than 0.29 particles per square centimeter (cm^2), with respect to particles having a diameter greater than 0.12 μm. This contamination level is equivalent to less than 90 localized light-scattering (LLS) defects greater than 0.12 μm on a 200 millimeter (mm) wafer. The roughness of top surface 32 may be less than 1 Å RMS in a 1 μm × 1 μm scan area. Further, top surface 32 of semiconductor layer 16 may have the following roughness and contamination levels:

Roughness root-mean-square	contamination level
< 5 Å (40 μm × 40 μm scan area)	<0.16 particles/cm^2
< 1 Å (1 μm × 1 μm scan area)	particle diameter > 0.16 μm
	(<50 LLS defects on a 200 mm wafer)
<5 Å (40 μm × 40 μm scan area)	<0.08 particles/cm^2
< 1 Å (1 μm × 1 μm scan area)	particle diameter > 0.2 μm
	(<25 LLS defects on a 200 mm wafer)

<5 Å (40 μm × 40 μm scan area) < 1 Å (1 μm × 1 μm scan area)	<0.019 particles/cm^2 particle diameter > 1.0 μm (<6 LLS defects on a 200 mm wafer)
<3 Å (40 μm × 40 μm scan area) <0.5 Å (1 μm × 1 μm scan area)	<0.09 particles/cm^2 particle diameter > 0.09 μm

The embodiments discussed above illustrate instances in which an annealing step helps eliminate superlattices, thereby reducing surface roughness. In some embodiments, however, an anneal can help reduce haze and provide a smoother layer surface even for layers which are initially homogeneous, i.e., do not have superlattice or columnar compositional variations.

In some embodiments, growth conditions, including a first plurality of parameters may be selected to prevent compositional superlattice formation, thereby eliminating the need for the aforementioned anneal. The first plurality of parameters may include temperature, precursor, growth rate, and pressure. For example, a superlattice-free SiGe graded buffer layer may be grown at high temperatures under the following conditions:

System: ASM EPSILON ® 2000 epitaxial reactor, manufactured by ASM International B.V., based in Bilthoven, the Netherlands

Temperature: 1000 - 1100°C

Pressure: 20 Torr to 760 Torr (atmospheric pressure)

Hydrogen flow: 20 - 80 standard liters per minute (slm)

Dichlorosilane flow: 50 - 250 standard cubic centimeters per minute (sccm)

Germanium tetrachloride flow: 0 - 0.5 gram per minute

Growth rate: 380 - 980 nm/min

In a preferred embodiment, conditions for growth of a superlattice-free graded SiGe buffer layer may be as follows:

System: ASM EPSILON® 2000 epitaxial reactor

Temperature: 1100°C

Pressure: 80 Torr

Hydrogen flow: 40 slm

Dichlorosilane flow: 250 sccm

Germanium tetrachloride flow: 0 - 0.5 gram per minute (for up to 20% Ge)

Growth rate: 850 - 980 nm/min

The presence or absence of a superlattice in a regrowth layer, e.g., a SiGe layer, formed after the planarization step should also be considered. Such a superlattice may be detrimental to the electrical properties of the semiconductor layer grown on it, e.g., a strained Si layer. In some embodiments, regrowth may be performed without forming a superlattice structure. Factors that reduce variation in a gas phase depletion profile in, e.g., a SiGe deposition system (and therefore also reduce upstream-to-downstream SiGe compositional variations) tend to reduce a tendency to define a superlattice in the SiGe layer. These factors include, for example, reduced dichlorosilane (DCS) or equivalent Si precursor flow/growth rate, decreased temperature, and increased hydrogen flow rates. Conditions that produce a difference of less than 5%, and preferably less than 2%, in the Ge fraction between the upstream and downstream positions on a wafer having a diameter of 200 millimeters (mm) or less may produce superlattice-free growth. A wafer having a diameter larger than 200 mm, e.g., 300 mm or larger, may require even less difference in the Ge fraction to achieve superlattice-free growth, e.g., possibly less than 2% variation. The effect of the conditions may be measured by growing a wafer without rotation and measuring upstream and downstream positions on the wafer near the wafer edge (<10 mm from a wafer edge, preferably <5 mm from the wafer edge).

Like for the semiconductor layer, the regrowth layer may be formed substantially haze-free, and may include two elements, the two elements being distributed to define a compositional variation within the semiconductor layer. A second plurality of parameters may be used for forming the regrowth layer. These parameters may include a second temperature, with the first temperature used to make the semiconductor layer being higher than the second temperature. As an example, superlattice-free regrowth of SiGe layers may be achieved in an ASM EPSILON® 2000 epitaxial reactor under the following representative conditions:

Temperature: 700 - 850°C

Pressure: 20 - 80 Torr

H_2 flow: 20 - 80 slm

Dichlorosilane flow: 20 - 60 sccm

Germane (GeH_4) flow: 8 - 34 sccm of 25% GeH_4

Growth rate: 20 - 200 nm/min

In a preferred embodiment, conditions for superlattice-free regrowth of SiGe layers may be as follows:

Temperature: 750 - 800°C

Pressure: 80 Torr

H_2 flow: 40 - 80 slm

Dichlorosilane flow: 50 sccm

Germane flow: 17 - 34 sccm of 25% GeH_4

Growth rate: 90 - 100 nm/min

Referring to Figure 10, in an alternative embodiment, the semiconductor layer 16 may have a lower portion 100 that includes a superlattice and an upper portion 110 disposed over the lower portion 100 that is substantially free of a superlattice. The superlattice of the lower portion 100 may help block the effects of an underlying misfit array, thereby enabling the suppression of the reappearance of cross-hatch during subsequent regrowth or post-planarization anneal steps.

Referring to Figure 11, in an embodiment, second layer 50 disposed over semiconductor layer 16 may be a regrowth layer having a lower portion 150 that includes a superlattice and an upper portion 152 that is substantially free of a superlattice. Performing the initial portion of the regrowth under conditions that promote the presence of the superlattice may block strain fields from an underlying misfit array. This enables suppression of the reappearance of the cross-hatch during the regrowth process or post-planarization anneal steps. A final portion of the regrowth can be performed using superlattice-free conditions as outlined above such that the final strained Si device layer is not in proximal contact with a region having a superlattice.

In some embodiments, a "buried" region may have a superlattice, e.g., lower portion 150 of regrowth second layer 50 or lower portion 100 of semiconductor layer 16, that may be annealed away after completion of epitaxial steps.

Illustrative Embodiments

Experimental Set 1

The following two SiGe relaxed buffer layer samples were analyzed with and without annealing:

1. Sample A: Non-annealed test wafer subjected to x-ray diffraction (XRD) measurement. The Ge composition was determined to be 29.5 ± 0.3%, with relaxation of 95.5 ± 1%.

2. Sample B: wafer whose Ge content was made more uniform by annealing. After deposition, the wafer was annealed in the same deposition chamber at 1050 °C for 5 minutes.

AFM analysis was conducted for samples A and B at different scan sizes (1 μm × 1 μm, 10 μm × 10 μm, and 50 μm × 50 μm). Referring to Table 1, roughness values [RMS and R_a (average roughness)] were obtained. Surface roughness increased by an average of about 20% after annealing, based on large scan sizes, i.e., 10 μm × 10 μm and 50 μm × 50 μm. Scans of a given size can capture roughness with wavelengths less than the scan size, but not larger. However, characteristic RMS values represent only the wavelength with the largest amplitude, i.e., the long wavelength. The layers in samples A and B do not exhibit columnar decomposition. Cross-hatch roughness, i.e., waviness, increases because of the thermal annealing of the sample. The cross-hatch does not correspond to the columnar decomposition; rather, it ultimately arises from the influence of the strain fields of the buried misfit dislocations in the graded layer. Annealing may cause the cross-hatch to reappear even after the layer has been polished because the surface atom mobility may be high at high temperatures. Because the buried misfit dislocations are still present below the surface, the atoms on the surface may start to rearrange under the influence of the misfit dislocation strain fields, bringing back a milder version of the original cross-hatch. On the other hand, based on the small scan size that captures the short wavelength roughness (< 1 μm), the short wavelength roughness decreased by a factor of approximately seven. This significant reduction in the short wavelength roughness reduces the haze level observed on wafers annealed like sample B.

In some cases, annealing may reduce the short wavelength roughness and the associated haze level of a layer, but may increase the large wavelength roughness (e.g., the cross-hatch roughness). Therefore, it may be advantageous to perform the annealing step prior to planarization. In this manner, the anneal reduces the propensity of the short wavelength roughness to reappear in subsequent processing steps, and the planarization step reduces any long wavelength roughness that reappeared during annealing. Because the re-emergence of the long wavelength roughness results from high surface atom mobility and from atoms responding to underlying strain fields below the surface, low long scale roughness may be maintained during the annealing step in other ways. In order to reduce the surface mobility of atoms in a layer, the layer may be capped by a protective layer. This protective layer may include material that will not react with the surface being protected and that is easily removed selectively to the underlying

surface. Suitable material for the protective layer may be, for example, silicon dioxide (SiO_2) or silicon nitride (Si_3N_4). The presence of the protective layer decreases the mobility of atoms in the layer to be annealed, since the atoms no longer lie atop a free surface. Thus, if additional planarization is not desirable after the anneal, protective layers may be utilized to prevent the re-emergence of long wavelength surface roughness.

Table 1 Roughness of samples A and B at different scan sizes

Sample ID	1×1 μm scan		10×10 μm scan		50×50 μm scan	
	RMS (nm)	R_a (nm)	RMS (nm)	R_a (nm)	RMS (nm)	R_a (nm)
A (not annealed)	0.700	0.588	0.956	0.774	2.622	1.974
B (annealed)	0.103	0.083	1.151	0.945	3.471	2.213

<u>Experimental set 2</u>

In second experiment, a SiGe graded buffer layer grown at > 850°C was annealed at 1050°C for 5 minutes at atmospheric pressure in hydrogen. Before and after the annealing, the surface roughness was measured by AFM with different scan sizes (1 × 1 μm, 10 × 10 μm, and 50 × 50 μm) at the center, mid-radius, and edge of the wafer. In addition, haze measurements using a laser defect scanner (SURFSCAN 6220, available from KLA-Tencor) were compared between equivalent buffer layers, one unannealed and the other annealed. Referring to Table 2, the short spatial wavelength surface roughness derived from the 1 μm × 1 μm scan decreased after the anneal by an average of about 50%. AFM images (50 μm × 50 μm, 10 μm × 10 μm, and 1 μm × 1 μm) at the edge of the wafer were compared before and after anneal. The number of periods in the cross-hatch roughness decreased after the anneal.

Table 2 Roughness of samples A and B at different scan sizes

Scan dimension & position	Pre-anneal		Post-anneal	
	R_a	RMS	R_a	RMS
50 μm - edge	5.766	4.612	5.322	7.047
10 μm - edge	4.119	3.318	2.643	3.320

1 μm - edge	0.730 0.434 0.508	1.463 0.543 0.640	0.210 0.383 0.274	0.266 0.922 0.300
50 μm - mid-radius	5.588	4.560	3.942	4.950
10 μm - mid-radius	3.446	2.839	2.964	4.041
1 μm - mid-radius	0.574	0.454	0.274	0.340
50 μm - center	6.189	4.957	3.641	4.689
10 μm - center			2.964	3.490
1 μm - center	0.669	0.584	0.257	0.311

<u>Laser Particle Scanner – Haze</u>

Surface roughness has a significant impact on the characterization of the buffer layers by laser particle scanning, e.g., with a Tencor SURFSCAN 6220. Higher roughness is observed as elevated haze levels, making detection of small particles difficult. For this reason, one of the key measurements indicating the effect of a process is the measurement of haze levels on the wafers.

Haze level measurements were made before and after the anneal of wafers having equivalent buffer layers. The haze levels of non-annealed and annealed wafers were compared, with wafers placed in the inspection tool in the "notch down" (0 degree rotation) orientation. Haze is measured as a fraction of light energy scattered by the surface relative to the energy in the incident laser beam. The haze level was reduced by 50% or more by the anneal, confirming the reduction of small scale roughness shown in the AFM data.

Another aspect of the effect of the anneal process on the wafer surface roughness and resulting haze measurement is the greater extent to which the haze of an annealed substrate is reduced by changing the orientation angle. Because fine scale roughness has a more random orientation than cross-hatch, the scattering characteristics of fine scale roughness do not depend on the orientation of the wafer in relation to the incident beam. The cross-hatch, in contrast, scatters the incident beam in a different direction depending on the orientation angle of the wafer.

Annealing the substrate increases the impact of the orientation angle on haze. Before an anneal, changing the orientation angle of the wafer (0 to 45 degrees) in the inspection system

reduces the measured haze by only about 10%, e.g., the average haze measurement is reduced from 716 to 657 ppm. After the anneal, the random, fine scale roughness is reduced, and the haze is reduced by 50% when the orientation angle is changed from 0 to 45 degrees.

Reduction of Vertical Superlattice Structure

A vertical superlattice, i.e., a vertical variation in the composition of the SiGe, has been observed in SiGe buffer layers.

X-ray diffraction (XRD) scans of buffer layer 14 provided evidence of the presence or absence of superlattices in buffer layers before and after anneal. XRD rocking curves were generated of a SiGe buffer layer 14 without an anneal and with an anneal for 1050°C for 5 minutes. Satellite peaks around the normal graded buffer signature (peaks at –3500 to –3000 arc-sec, and at +700 to +1000 arc-sec indicated the presence of the superlattice structure in buffer layer 14 that has not been annealed. The peaks were observed at the wafer edge, possibly due to the wafer edge alternating as a leading and trailing edge due to wafer rotation in a horizontal flow reactor. The satellite peaks were not present, neither at the center nor at the edge of the wafer, in a SiGe graded buffer layer 14 that has been annealed.

The invention may be embodied in other specific forms without departing from the spirit or essential characteristics thereof. The foregoing embodiments are therefore to be considered in all respects illustrative rather than limiting on the invention described herein. Scope of the invention is thus indicated by the appended claims rather than by the foregoing description, and all changes which come within the meaning and range of equivalency of the claims are intended to be embraced therein.

What is claimed is: